## Simulation study plan

The ADEMP (Aims, Data-generating mechanisms, Estimands, Methods, Performance measures) framework has been used to design this simulation study, as recommended by Morris et al.[1]

### Aims

To investigate the performance of four statistical methods used to analyse longitudinal quality of life scores derived from the EORTC QLQ-C30 questionnaire.

### Data-generating mechanisms

Three simulation studies will be performed: the first will be a two-group parallel randomised controlled trial (RCT) where both groups receive a medical intervention at the start of the trial; the second will be a two-group parallel RCT where both groups receive a surgical intervention at the start of the trial; and the third will be a two-group parallel RCT where one group receives a surgical intervention at the start of the trial and one group receives a medical intervention at the start of the trial. Quality of life (QoL) and survival data from the VIOLET (ref 13/04/03) and MARS2 (ref 15/188/31) trials have been used to inform the parameter choices for the simulation studies detailed below, such as the distribution of the baseline QoL score. Survival data from specific groups from the two trials were used to estimate realistic baseline log hazard functions (e.g., the no surgery group from MARS2 was used to inform the baseline log hazard for the study simulating an RCT with a medical intervention in both groups). QoL data were used to inform covariance structures, but these were simplified to not over-complicate the simulation studies. Different treatment effects have been chosen to represent small and moderate treatment effects, including treatment effects changing with time, as seen in the VIOLET and MARS2 trials as well as other published RCTs.

The data-generating mechanisms (DGM) for each simulation study are defined as follows:

| **DGM for simulation study 1 (RCT with medical intervention)** | |
|---|---|
| Outcome of interest | Global health status/QoL (GHS), continuous outcome (scores range from 0-100) |
| Allocation ratio | 1:1, treatment group generated as a random variable from a Bernoulli distribution with probability 0.5 |
| Number of post-randomisation time points | 4 |
| Baseline score | Generated as a random variable from a Beta(a, b) distribution with mean ~70, SD 21 in both groups. As the GHS score can only take set values, the score generated will be assigned the closest score from the list of possible GHS scores. |
| Total sample size [*n*] | Two sample sizes: n=150, n=500 |
| Follow up time points [*F*] | Two follow up patterns: equally spaced time points at 6, 12, 18 and 24 months; unequally spaced time points with increasing distance between time points at 2, 6, 12 and 24 months |
| Global health status/QoL | GHS score will be generated from the following random intercept models, where equation (1) will be used to generate the GHS score when the true treatment effect is constant over time, and equation (2) will be used to generate the GHS score when the true treatment effect changes over time. Generated scores will be assigned the closest score from the list of possible GHS scores.<br><br>$$GHS_{ij} = \alpha + \beta_1 baseline\ GHS_i + \beta_2 group_i + \beta_3 time_{ij} + u_i + \varepsilon_{ij}$$ (1) |

| | |
|---|---|
| | *where* $\boldsymbol{i}$ *denotes patient,* $\boldsymbol{j}$ *denotes time,* $\boldsymbol{\alpha = 28, \beta_1 = 0.5}$*,* $\boldsymbol{\beta_2}$ $\boldsymbol{and}$ $\boldsymbol{\beta_3}$ *are matrices corresponding to the treatment effects described below,* $\boldsymbol{u_i \sim N(0, 2)}$ *and* $\boldsymbol{\varepsilon_{ij}}$ *is drawn from a multivariate normal distribution with mean 0 and covariance structure as defined below.*<br><br>$\boldsymbol{GHS_{ij} = \alpha + \beta_1 baseline\ GHS_i + \beta_2 group_i + \beta_3 time_{ij} + \beta_3 group * time + u_i + \varepsilon_{ij}}$ (2) |
| Covariance structure | Toeplitz covariance matrix:<br>$= \begin{pmatrix} \mathbf{1} & \mathbf{0.5} & \mathbf{0.4} & \mathbf{0.35} \\ \mathbf{0.5} & \mathbf{1} & \mathbf{0.5} & \mathbf{0.4} \\ \mathbf{0.4} & \mathbf{0.5} & \mathbf{1} & \mathbf{0.5} \\ \mathbf{0.35} & \mathbf{0.4} & \mathbf{0.5} & \mathbf{1} \end{pmatrix}$ |
| True treatment effect (mean difference (MD)) [$\beta$] | *Three treatment effects for equally spaced follow up time points:*<br>1) MD constant over time, small treatment effect (MD=3)<br>Mean group 1 (68, 69, 71, 73); mean group 2 (71, 72, 74, 76)<br><br>2) MD constant over time, large treatment effect (MD=10)<br>Mean group 1 (64, 65, 66, 68); mean group 2 (74, 75, 76, 78)<br><br>3) MD changes over time (MD=7, 13, 10, 4)<br>Mean group 1 (67, 63, 66, 71); mean group 2 (74, 76, 76, 75)<br><br>*Three treatment effects for unequally spaced follow up time points:*<br>1) MD constant over time, small treatment effect (MD=3)<br>Mean group 1 (69, 71, 72, 73); mean group 2 (72, 74, 75, 76)<br><br>2) MD constant over time, large treatment effect (MD=10)<br>Mean group 1 (65, 66, 67, 68); mean group 2 (75, 76, 77, 78)<br><br>3) MD changes over time (MD=7, 13, 10, 4)<br>Mean group 1 (66, 62, 66, 72); mean group 2 (73, 75, 76, 76) |
| Missing data due to death [*D*] | Base case: no deaths<br><br>Survival times will be simulated and used to generate missing data due to death. The baseline log hazard function will be modelled using a fractional polynomial with one turning point (FP2 (0, 2)) defined as:<br>$\boldsymbol{\log(h_0(t)) = -1.8 - 0.15t^2 + 0.31log(t)}$<br><br>Survival times will be simulated from this underlying baseline hazard function using a flexible parametric survival model, with treatment group as a binary covariate with an associated log hazard ratio (HR).[2, 3]<br><br>Two log HRs for treatment group: log HR = -0.36 (HR=0.7); log HR=-0.02 (HR=0.98). Administrative censoring at 24 months will be applied. |
| Data missing at random (MAR) [*M*] | Base case: no missing data<br><br>Data will be simulated to be MAR. No baseline data will be set to missing.<br><br>1) Data will be set to missing if the GHS score from the previous time point is <50 with a probability of 10% (a lower GHS score indicates a lower QoL/worse health). |

| | |
|---|---|
| | 2) Data will be set to missing if the GHS score from the previous time point is <50 with a probability of 40% |

| **DGM for simulation study 2 (RCT with surgical intervention)** | |
|---|---|
| Outcome of interest | GHS, continuous outcome (scores range from 0-100) |
| Allocation ratio | 1:1, treatment group generated as a random variable from a Bernoulli distribution with probability 0.5 |
| Number of post-randomisation time points | 4 |
| Baseline score | Generated as a random variable from a Beta(a, b) distribution with mean ~70, SD ~21 in both groups. As the GHS score can only take set values, the score generated will be assigned the closest score from the list of possible GHS scores. |
| Total sample size [*n*] | Two sample sizes: n=150, n=500 |
| Follow up time points [*F*] | Two follow up patterns: equally spaced time points at 6, 12, 18 and 24 months; unequally spaced time points with increasing distance between time points at 2, 6, 12 and 24 months |
| Global health status/QoL | GHS score will be generated from the following random intercept models, where equation (1) will be used to generate the GHS score when the true treatment effect is constant over time, and equation (2) will be used to generate the GHS score when the true treatment effect changes over time. Generated scores will be assigned the closest score from the list of possible GHS scores.<br><br>$\boldsymbol{GHS_{ij} = \alpha + \beta_1 baseline\ GHS_i + \beta_2 group_i + \beta_3 time_{ij} + u_i + \varepsilon_{ij}}$ (1)<br><br>*where* $\boldsymbol{i}$ *denotes patient,* $\boldsymbol{j}$ *denotes time,* $\boldsymbol{\alpha = 28, \beta_1 = 0.5}$*,* $\boldsymbol{\beta_2\ and\ \beta_3}$ *are matrices corresponding to the treatment effects described below,* $\boldsymbol{u_i \sim N(0, 2)}$ *and* $\boldsymbol{\varepsilon_{ij}}$ *is drawn from a multivariate normal distribution with mean 0 and covariance structure as defined below.*<br><br>$\boldsymbol{GHS_{ij} = \alpha + \beta_1 baseline\ GHS_i + \beta_2 group_i + \beta_3 time_{ij} + \beta_3 group * time + u_i + \varepsilon_{ij}}$ (2) |
| Covariance structure | Toeplitz covariance matrix:<br>$$= \begin{pmatrix} 1 & 0.5 & 0.4 & 0.35 \\ 0.5 & 1 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1 & 0.5 \\ 0.35 & 0.4 & 0.5 & 1 \end{pmatrix}$$ |
| True treatment effect (MD) [$\beta$] | *Three treatment effects for equally spaced follow up time points:*<br>1) MD constant over time, small treatment effect (MD=3)<br>Mean group 1 (60, 66, 67, 69); mean group 2 (63, 69, 70, 72)<br><br>2) MD constant over time, large treatment effect (MD=10)<br>Mean group 1 (55, 61, 61, 63); mean group 2 (65, 71, 71, 73)<br><br>3) MD changes over time (MD=12, 9, 5, -2)<br>Mean group 1 (51, 57, 67, 74); mean group 2 (63, 66, 72, 72)<br><br>*Three treatment effects for unequally spaced follow up time points:*<br>1) MD constant over time, small treatment effect (MD=3)<br>Mean group 1 (55, 60, 66, 69); mean group 2 (58, 63, 69, 72) |

| | |
|---|---|
| | 2) MD constant over time, large treatment effect (MD=10)<br>Mean group 1 (50, 55, 61, 63); mean group 2 (60, 65, 71, 73)<br><br>3) MD changes over time (MD=12, 9, 5, -2)<br>Mean group 1 (50, 53, 60, 74); mean group 2 (62, 62, 65, 72) |
| Missing data due to death [*D*] | Base case: no deaths<br><br>Survival times will be simulated and used to generate missing data due to death. The baseline log hazard function will be modelled using a fractional polynomial with two turning points (FP3 (0.5, 3, 3)) defined as:<br>$\mathbf{log}(\boldsymbol{h_0(t)}) = -\mathbf{2.9} + \mathbf{1.3}\boldsymbol{t}^{\mathbf{0.5}} - \mathbf{0.12}\boldsymbol{t}^{\mathbf{3}} - \mathbf{0.66}\boldsymbol{t}^{\mathbf{3}}\boldsymbol{log(t)}$<br><br>Survival times will be simulated from this underlying baseline hazard function using a flexible parametric survival model, with treatment group as a binary covariate with an associated log hazard ratio HR.<br><br>Two log HRs for treatment group: log HR = -0.36 (HR=0.7); log HR=-0.02 (HR=0.98). Administrative censoring at 24 months will be applied. |
| Data MAR [*M*] | Base case: no missing data<br><br>Data will be simulated to be MAR. No baseline data will be set to missing.<br><br>1) Data will be set to missing if the GHS score from the previous time point is <50 with a probability of 10%<br>2) Data will be set to missing if the GHS score from the previous time point is <50 with a probability of 40% |

| **DGM for simulation study 3 (RCT with medical and surgical interventions)** | |
|---|---|
| Outcome of interest | GHS, continuous outcome (scores range from 0-100) |
| Allocation ratio | 1:1, treatment group generated as a random variable from a Bernoulli distribution with probability 0.5 |
| Number of post-randomisation time points | 4 |
| Baseline score | Generated as a random variable from a Beta(a, b) distribution with mean ~70, SD ~21 in both groups. As the GHS score can only take set values, the score generated will be assigned the closest score from the list of possible GHS scores. |
| Total sample size [*n*] | Two sample sizes: n=150, n=500 |
| Follow up time points [*F*] | Two follow up patterns: equally spaced time points at 6, 12, 18 and 24 months; unequally spaced time points with increasing distance between time points at 2, 6, 12 and 24 months |
| Global health status/QoL | GHS score will be generated from the following random intercept models, where equation (1) will be used to generate the GHS score when the true treatment effect is constant over time, and equation (2) will be used to generate the GHS score when the true treatment effect changes over time. Generated scores will be assigned the closest score from the list of possible GHS scores.<br><br>$\boldsymbol{GHS_{ij}} = \boldsymbol{\alpha} + \boldsymbol{\beta_1 baseline\ GHS_i} + \boldsymbol{\beta_2 group_i} + \boldsymbol{\beta_3 time_{ij}} + \boldsymbol{u_i} + \boldsymbol{\varepsilon_{ij}}$ (1) |

| | |
|---|---|
| | *where* $\boldsymbol{i}$ *denotes patient,* $\boldsymbol{j}$ *denotes time,* $\boldsymbol{\alpha = 28, \beta_1 = 0.5}$*,* $\boldsymbol{\beta_2}$ $\boldsymbol{and}$ $\boldsymbol{\beta_3}$ *are matrices corresponding to the treatment effects described below,* $\boldsymbol{u_i \sim N(0, 2)}$ *and* $\boldsymbol{\varepsilon_{ij}}$ *is drawn from a multivariate normal distribution with mean 0 and covariance structure as defined below.*<br><br>$\boldsymbol{GHS_{ij} = \alpha + \beta_1 baseline\ GHS_i + \beta_2 group_i + \beta_3 time_{ij} + \beta_3 group * time + u_i + \varepsilon_{ij}}$ (2) |
| Covariance structure | Toeplitz covariance matrix:<br>$= \begin{pmatrix} \mathbf{1} & \mathbf{0.5} & \mathbf{0.4} & \mathbf{0.35} \\ \mathbf{0.5} & \mathbf{1} & \mathbf{0.5} & \mathbf{0.4} \\ \mathbf{0.4} & \mathbf{0.5} & \mathbf{1} & \mathbf{0.5} \\ \mathbf{0.35} & \mathbf{0.4} & \mathbf{0.5} & \mathbf{1} \end{pmatrix}$ |
| True treatment effect (MD) [$\beta$] | *Three treatment effects for equally spaced follow up time points:*<br>1) MD constant over time, small treatment effect (MD=3)<br>Mean group 1 (64, 66, 68, 70); mean group 2 (67, 69, 71, 73)<br><br>2) MD constant over time, large treatment effect (MD=10)<br>Mean group 1 (53, 59, 61, 64); mean group 2 (63, 69, 71, 74)<br><br>3) MD changes over time (MD=14, 7, 2, -5)<br>Mean group 1 (53, 61, 66, 74); mean group 2 (67, 68, 68, 69)<br><br>*Three treatment effects for unequally spaced follow up time points:*<br>1) MD constant over time, small treatment effect (MD=3)<br>Mean group 1 (58, 64, 66, 70); mean group 2 (61, 67, 69, 73)<br><br>2) MD constant over time, large treatment effect (MD=10)<br>Mean group 1 (50, 53, 59, 64); mean group 2 (60, 63, 69, 74)<br><br>3) MD changes over time (MD=14, 7, 2, -5)<br>Mean group 1 (50, 57, 64, 74); mean group 2 (64, 64, 66, 69) |
| Missing data due to death [*D*] | Survival times will be simulated and used to generate missing data due to death. The baseline log hazard function will be modelled using a fractional polynomial with two turning point (FP3 (0, 3, 3)) defined as:<br>$\boldsymbol{\log(h_0(t)) = -1.6 + 0.5 log(t) - 0.06 t^3 - 1.03 t^3 \log(t)}$<br><br>Survival times will be simulated from this underlying baseline hazard function using a flexible parametric survival model, with treatment group as a binary covariate with an associated log HR.<br><br>Two log HRs for treatment group: log HR = -0.36 (HR=0.7); log HR=-0.02 (HR=0.98). Administrative censoring at 24 months will be applied. |
| Data MAR [*M*] | Base case: no missing data<br><br>Data will be simulated to be MAR. No baseline data will be set to missing.<br><br>1) Data will be set to missing if the GHS score from the previous time point is <50 with a probability of 10%<br>2) Data will be set to missing if the GHS score from the previous time point is <50 with a probability of 40% |

Number of scenarios (fully factorial design) and simulation runs:

- Each simulation study will investigate $\boldsymbol{n \times F \times \beta \times D \times M = 2 \times 2 \times 3 \times 3 \times 3}$ = 108 scenarios.
- N=5000 simulation repetitions per scenario. The number of simulation repetitions has been determined using the formula $\boldsymbol{n_{sim} = (\frac{Z_{1-(\alpha/2)}\sigma}{\delta})^2}$ where $\boldsymbol{Z_{1-(\alpha/2)}}$ is the $\boldsymbol{1-(\alpha/2)}$ quantile of the standard normal distribution, $\boldsymbol{\sigma^2}$ is the variance of the parameter of interest (MD), and $\delta$ is the specified level of accuracy of the estimate of interest we are willing to accept, as described by Burton *et al*.[4] The minimum effect size to be simulated is 2, and assuming a maximum standard deviation of the MD of 3.6 and a 5% significance level, then 5000 simulation repetitions will allow estimation of the MD within 5% accuracy of the true coefficient. The number of simulation repetitions will be increased if the Monte Carlo standard errors (MCSEs) of key performance measures are not acceptably small, (e.g. larger than 0.5% on a coverage of 95%) indicating further repetitions are required.

Moderately independent simulations will be generated, whereby the same set of simulated datasets will be used to compare the different statistical methods of interest for each scenario, but a different set of datasets will be generated for each scenario investigated.[4]

**Estimand**

Target estimand = mean difference between the two treatment groups

It will be assumed that adherence to randomised group is perfect, and all patients (excluding deaths) are followed up to the end of the study (24 months).

**Methods**

The four methods to be compared will include:

1) Linear mixed effects model
2) Joint longitudinal survival model
3) T-test
4) Generalised estimating equations

Analyses will be performed on an intention to treat basis.

**Performance measures**

The performance measures to be presented, along with their associated MCSE and/or Monte Carlo 95% confidence intervals, will include:

- Bias
- Mean square error
- Power
- Ratio of Model standard error/empirical standard error
- Coverage of confidence intervals

These performance measures have been chosen as we are interested in evaluating methods for estimating intervention effects.

**References**

1. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Statistics in Medicine. 2019;38(11):2074-102.
2. Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. Statistics in Medicine. 2013;32(23):4118-34.
3. Austin PC, Fang J, Lee DS. Using fractional polynomials and restricted cubic splines to model non-proportional hazards or time-varying covariate effects in the Cox regression model. Statistics in Medicine. 2022;41(3):612-24.
4. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Statistics in Medicine. 2006;25(24):4279-92.